The Particle Swarm: Social Interaction as Intelligence

NASA Goddard Space Flight Center
9/26/2012

James Kennedy
Washington, DC
Kennedy.Jim@gmail.com

The Particle Swarm

A stochastic, population-based algorithm for problem solving

Based on a social-psychological metaphor

Used by engineers, computer scientists, applied mathematicians, etc.

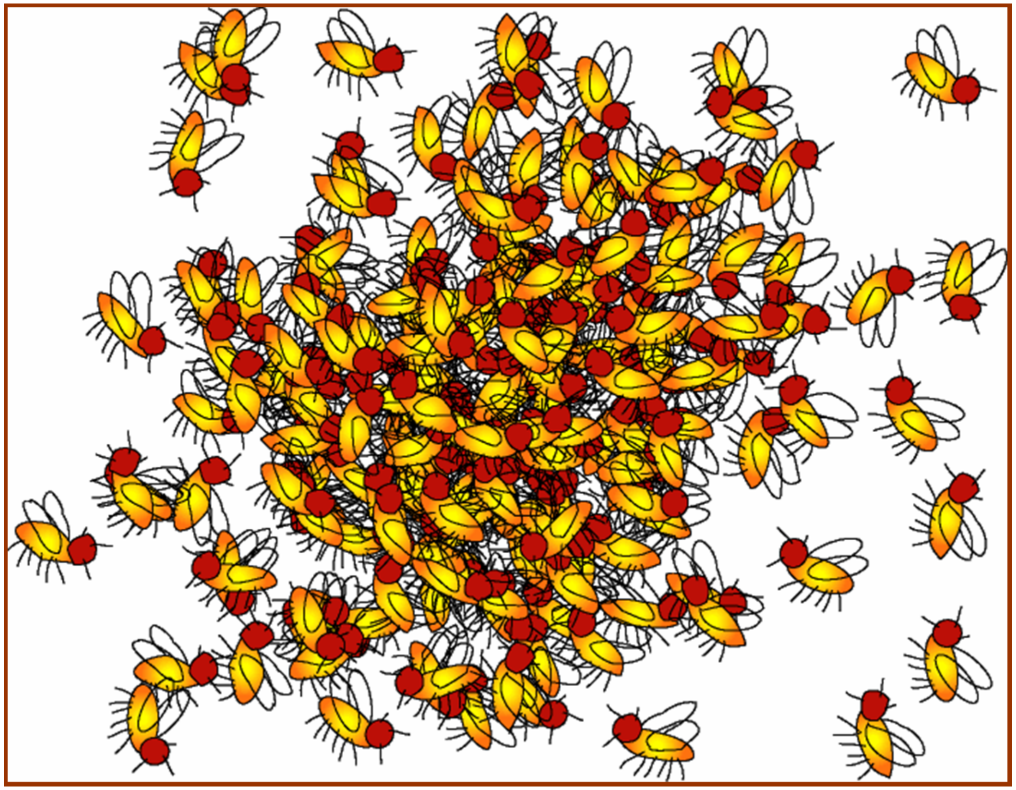
First reported in 1995 by Kennedy and Eberhart

Constantly evolving

The Particle Swarm Paradigm is a Particle Swarm

A population of individuals interact with one another according to simple rules in order to solve problems, which may be very complex.

It is an appropriate kind of description of the process of science.



Easy to implement in code, not easy to understand how it works.

Dynamic Social Impact Theory

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97*, 362-376.

i=f(SIN)

Computer simulation – 2-d CA

Each individual is both a target and source of influence

“Euclidean” neighborhoods

Binary, univariate individuals

“Strength” randomly assigned

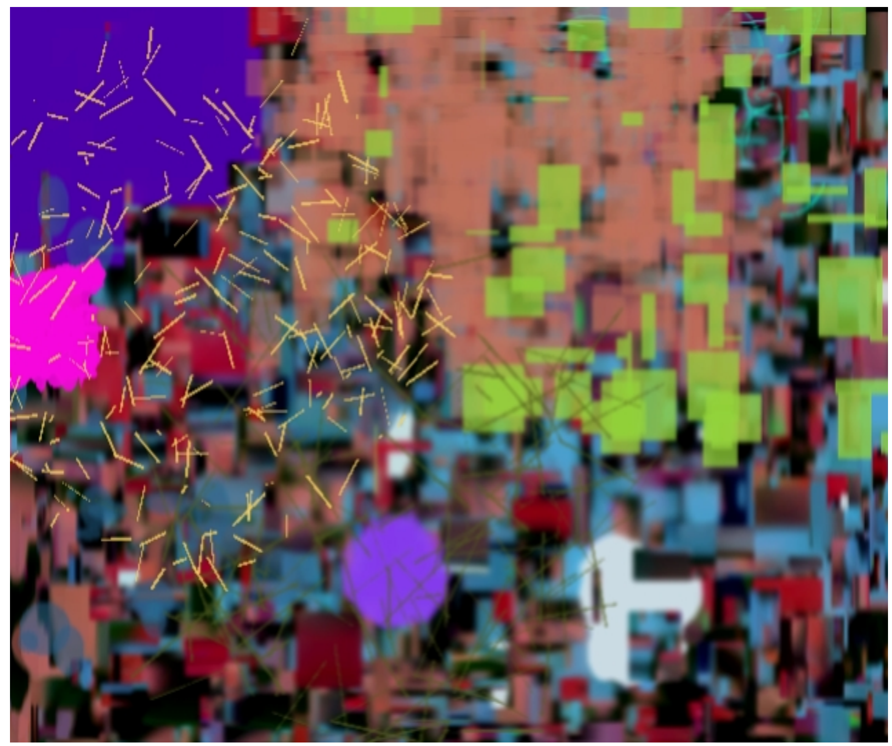
The result: Polarization

```
---------IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
---------IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
----------IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
-----------IIIIIIIIIIIIIIIIIIIIIIIIIIIII
-----------IIIIIIIIIIIIIIIIIIIIIIIIIIIII
----------IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
-------IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
----IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
I-IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIII-IIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIII-I-IIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIII----IIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIII----IIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIII----IIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII--
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII--
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII-II
IIIIIIIIIIIIIIIIIIIIIIIII-IIIIIIIIIII-II
IIIIIIII-----IIIIIIIIIIIII-IIIIIIIIIIIII
IIIII---------IIIIIIIIIIIIIIIII--IIIIII-
IIII-----------IIIIIIIIIIIIIII------IIII
```

Evolutionary Computation

Genetic algorithms
Evolutionary Programming
Evolution Strategies
Genetic programming

Population based stochastic methods
Mutation
Crossover
Iteration = generation



Birds, Fish – and Minds

Biological models

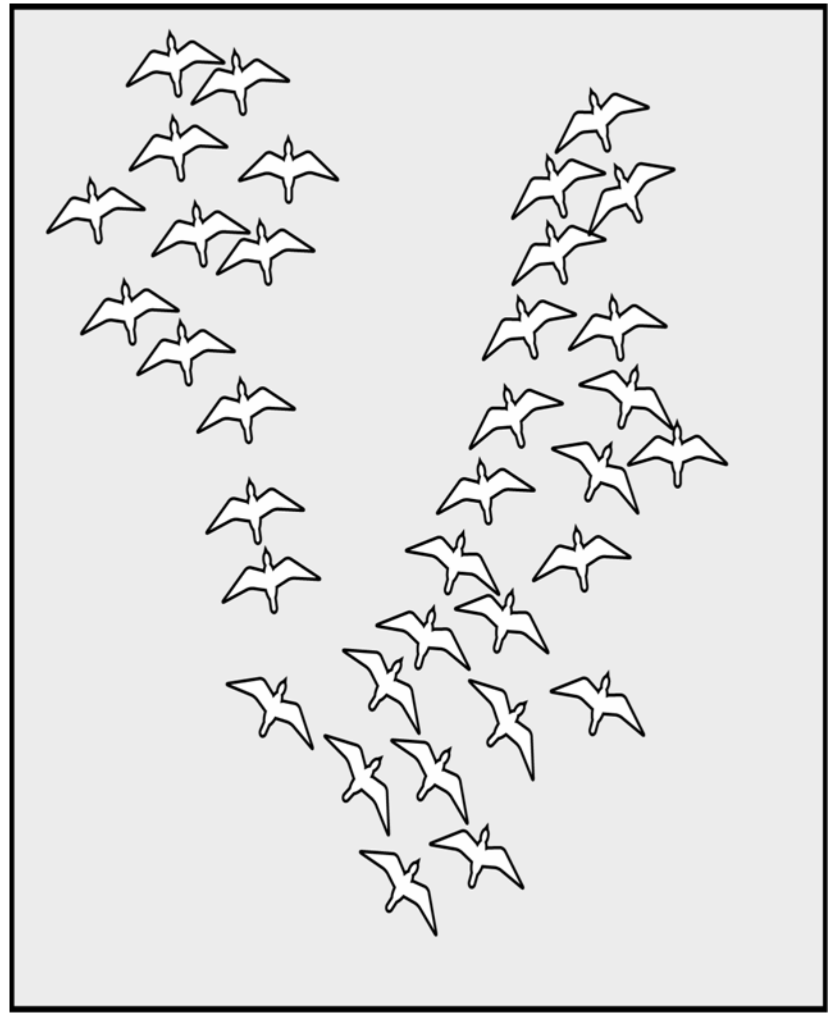
Simulations
Heppner & Grenander
Reynolds

Feed – birds notice that other birds notice

Euclidean neighborhoods

Thinking as cognitive search and optimization (cognitive dissonance)

Note: Collision / agreement

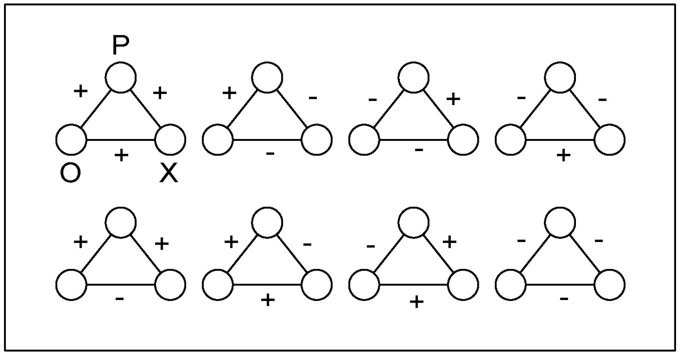


(BatchNet source code = bird feed)

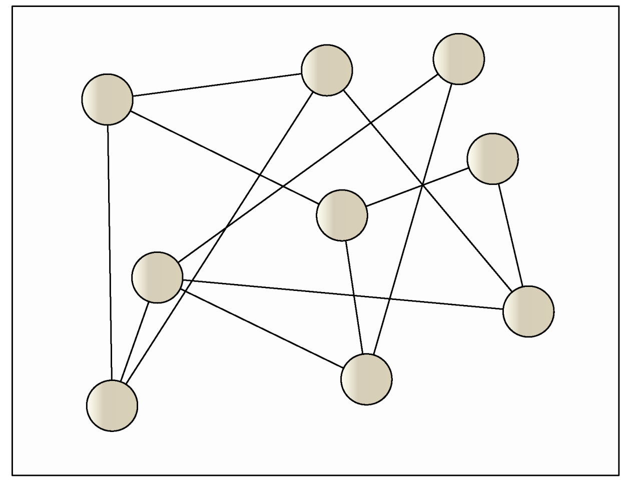
Cognition as Optimization

Minimizing or maximizing a function result by adjusting parameters

Cognitive consistency theories, incl. dissonance

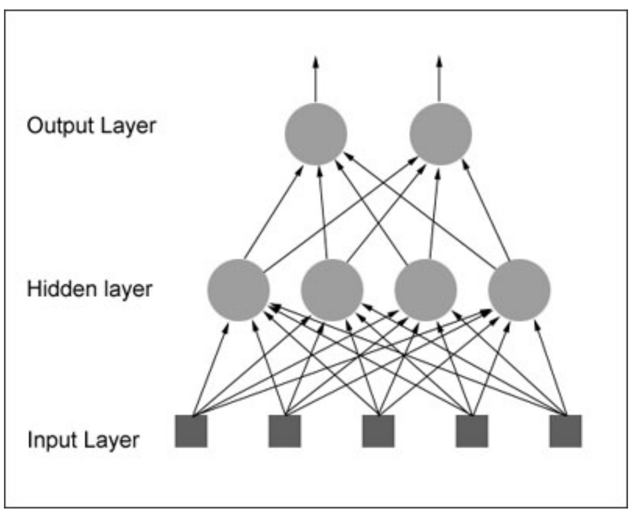


Parallel constraint satisfaction



Connectionism

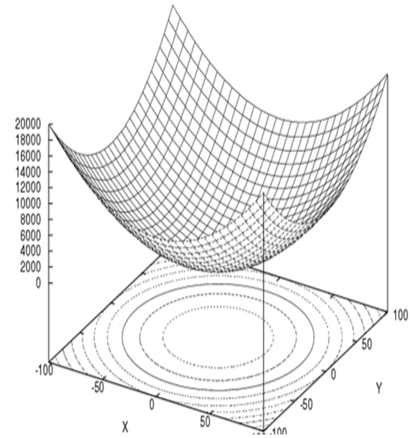
Feedforward Neural Nets



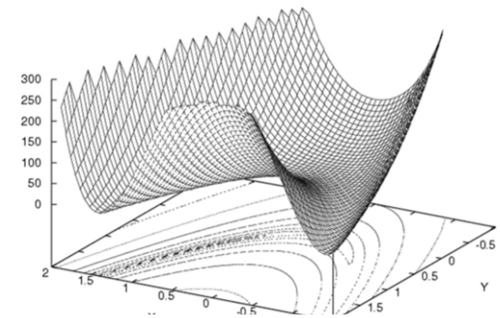
Particle swarm works with the dynamics of the network, as opposed to its equilibrium properties

Some Standard Test Functions

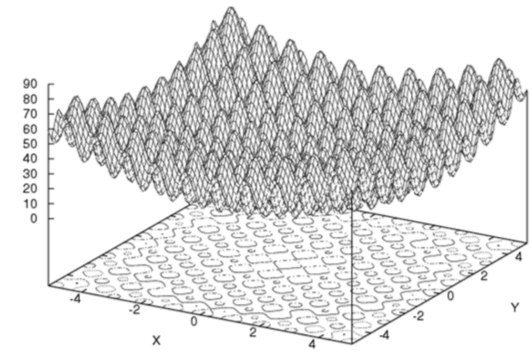
Sphere



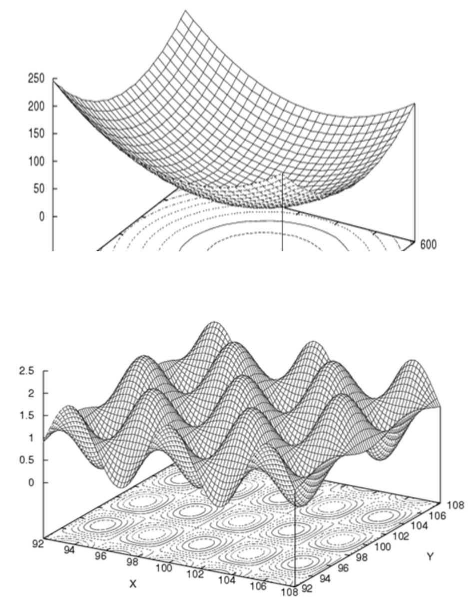
Rosenbrock



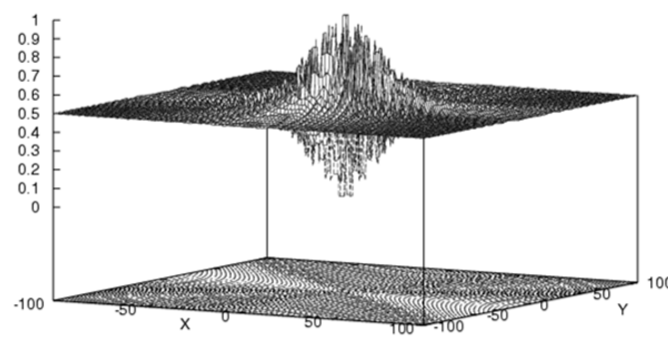
Rastrigin



Griewank



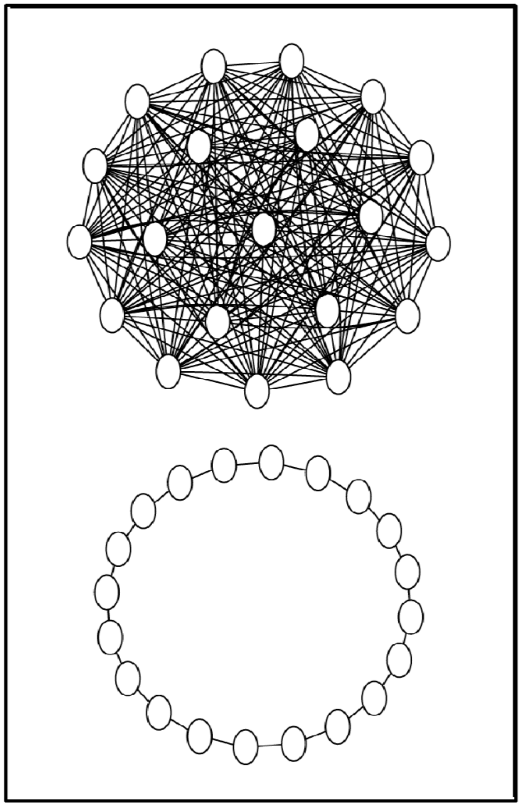
Schaffer’s f6



Classic PSO Fixed Communication Topologies

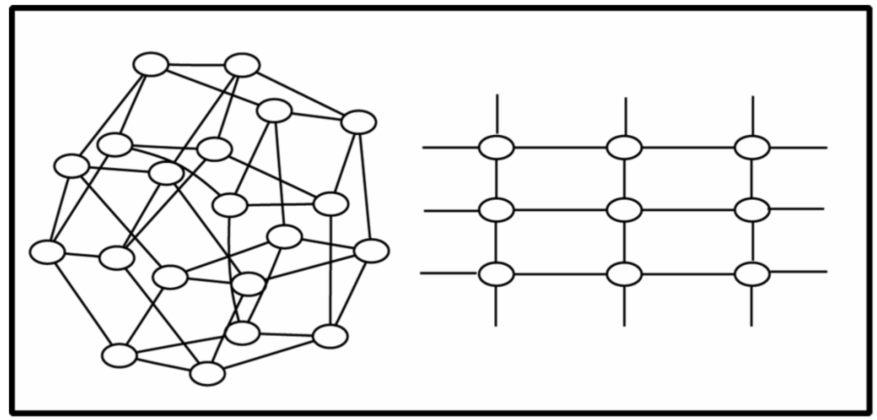
Gbest

Lbest



Since 1995

Von Neumann (Square) added later



Each particle is a search process. Each is a learner and a teacher simultaneously.

The Particle Swarm

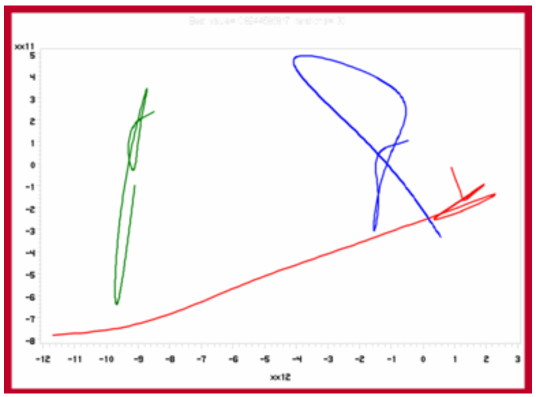
Initialize Population and constants
Repeat
Do i=1 to population size
$CurrentEval_i$ = eval($\vec{x}_i$)
If CurrentEval < $pbest_i$ then do
    $pbest_i$ = $CurrentEval_i$
    For d=1 to Dimension
        $p_{id} = x_{id}$
    Next d
    If $CurrentEval_i < Pbest_{gbest}$ then gbest=i
End if
g = best neighbor's index

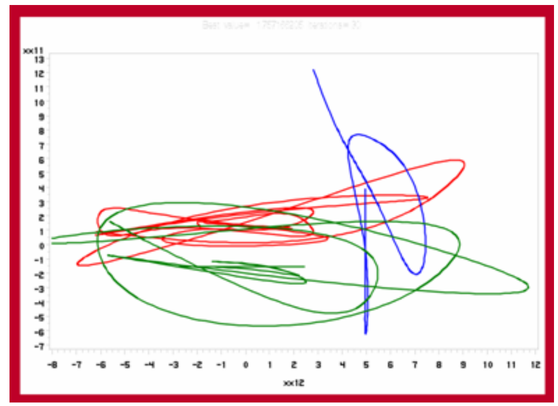
For d=1 to Dimension
    $v_{id} = W * v_{id} + U(0, AC) \times (p_{id} - x_{id}) + U(0, AC) \times (p_{gd} - x_{id})$
    $x_{id} = x_{id} + v_{id}$
Next d
Next i
Until termination criterion

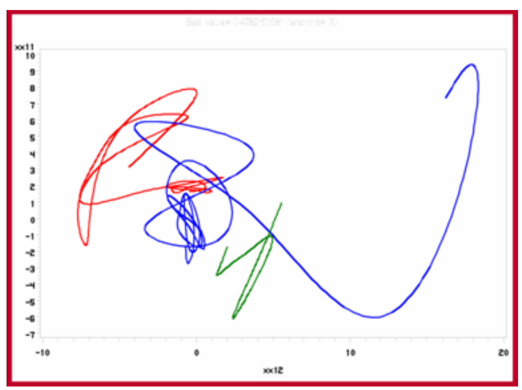
Without Interaction

No particle by itself is able to solve the problem

Communication is necessary

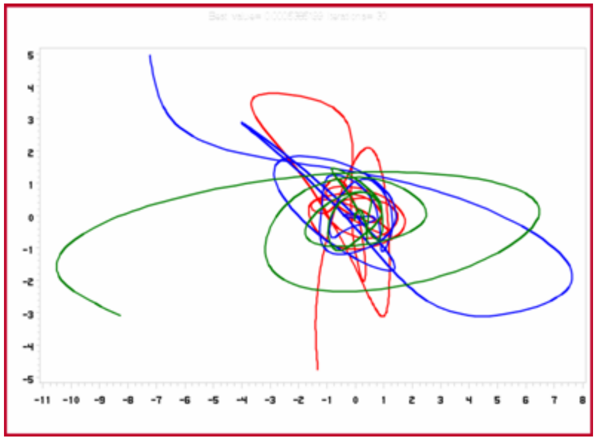


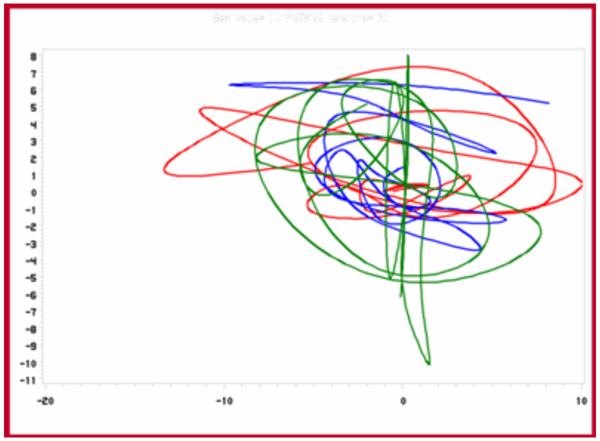


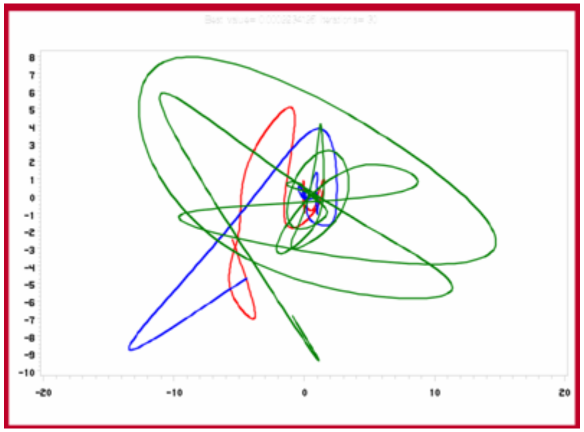


(2-D sphere function)

Collaboration





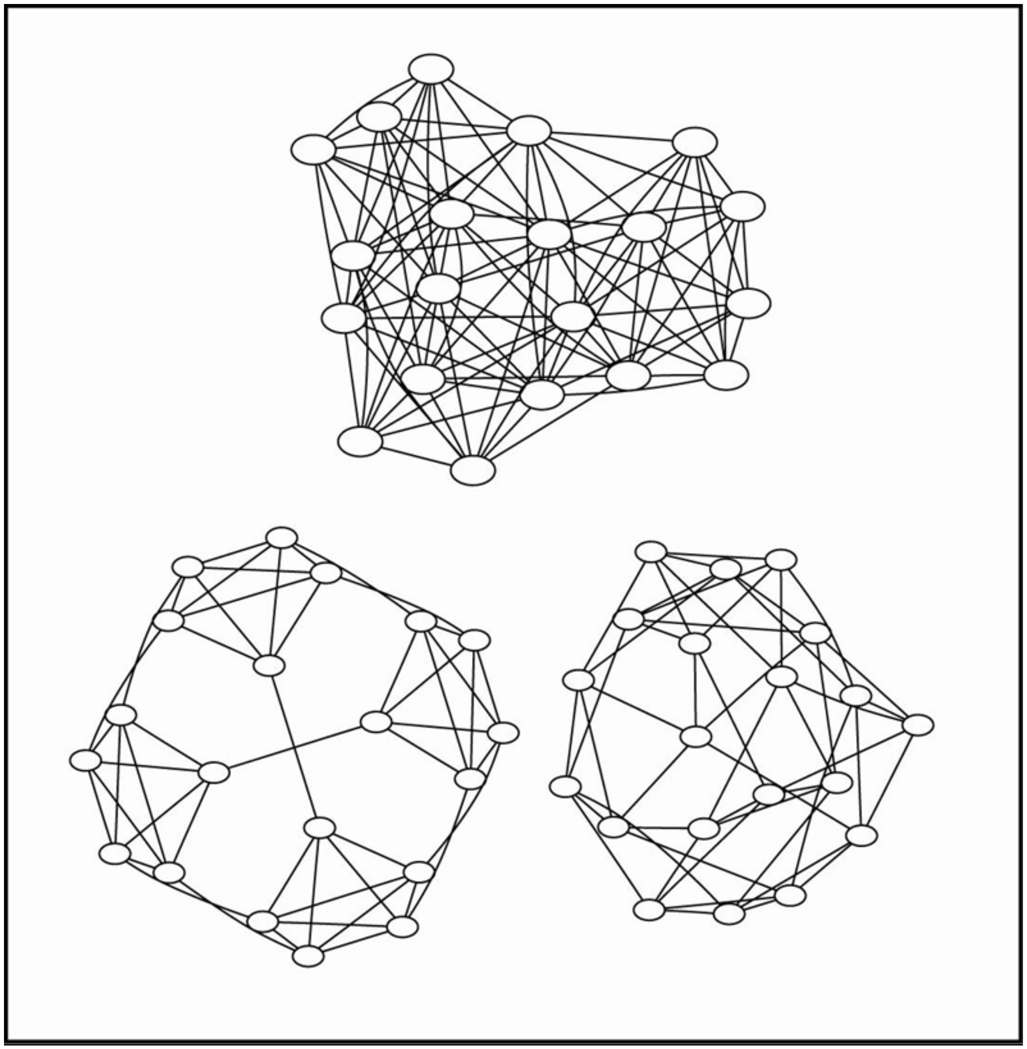


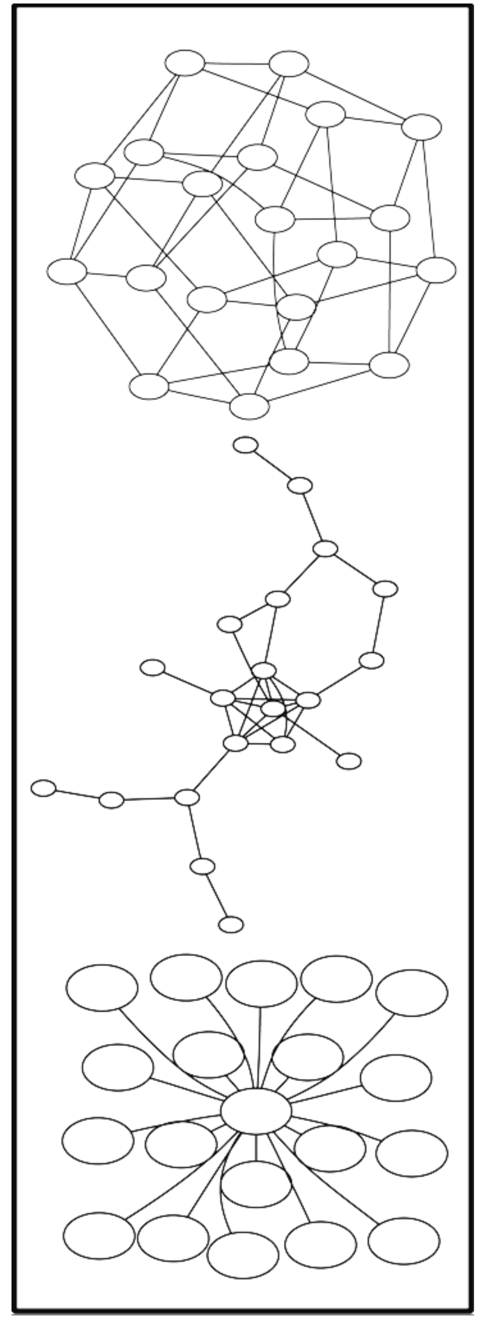
All work toward the same goal – minimize the same objective function

Clustering / converging near optimum

Some Topologies

What do you think will work?





Outliers, Misfits

Stochastic adaptive processes have costs

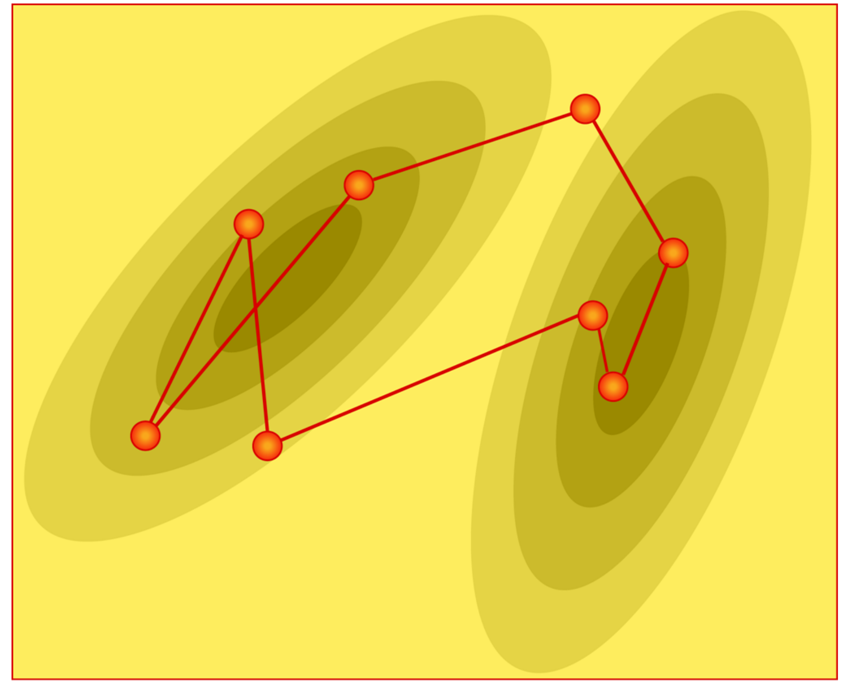
Redundancy, futile exploration

But …

They are robust, versatile, creative, and may even find solutions in "infeasible" space

System parameters determine the balance of exploration and exploitation

An individual attracted to multiple regions may find a new kind of solution, or may lead neighbors to better solutions they would not have found



Convergence and Clustering

Convergence: a numerical method is said to be convergent if the numerical solution approaches the exact solution as the step size h goes to 0.

Clustering: members of a population occupy proximal locations in the search space.

Seems incorrect: "When a population consists primarily of similar individuals, we say it has converged."

A particle swarm has both: individuals retain their identities over time

Convergence and clustering are correlated but different

Smaller steps as they gather in a region.

The Sociocognitive Principle

Correlations among three variables:

- Topological distance
- Euclidean distance
- Step size

Individuals who communicate more tend to become more similar (norm formation) – or conversely individuals who are more similar tend to communicate more (homophily)

As individuals' beliefs, attitudes, solution vectors, etc., become more similar, they tend to take smaller steps through the search space. Individuals become more conservative as groups become more uniform.

At beginning of a typical particle swarm trial, topological distances are assigned, position in search space and initial step size are random. Correlations develop over time.

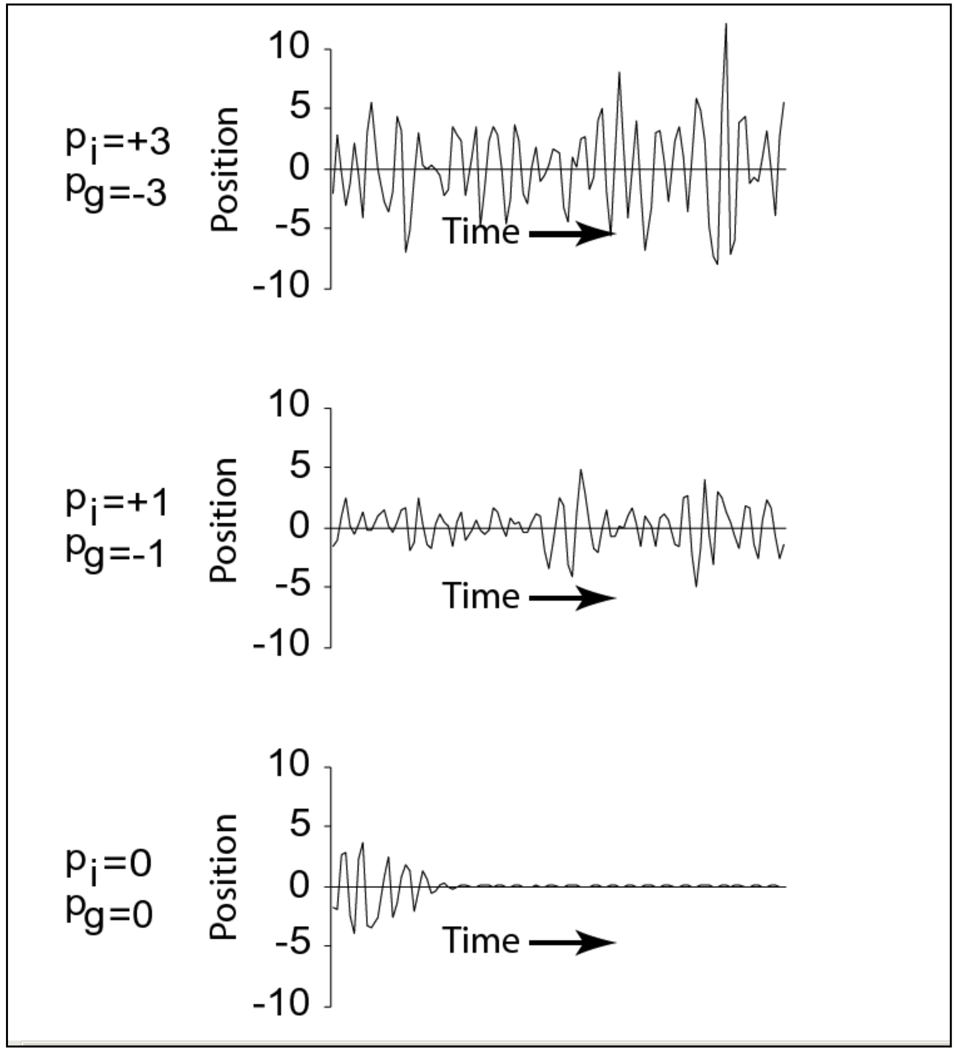
Step-Size Depends on Neighbors

Movement of the particle through the search space is centered on the midpoint between p_i and p_g on each dimension, and its amplitude is scaled to their difference.

Exploration vs. exploitation: automatic

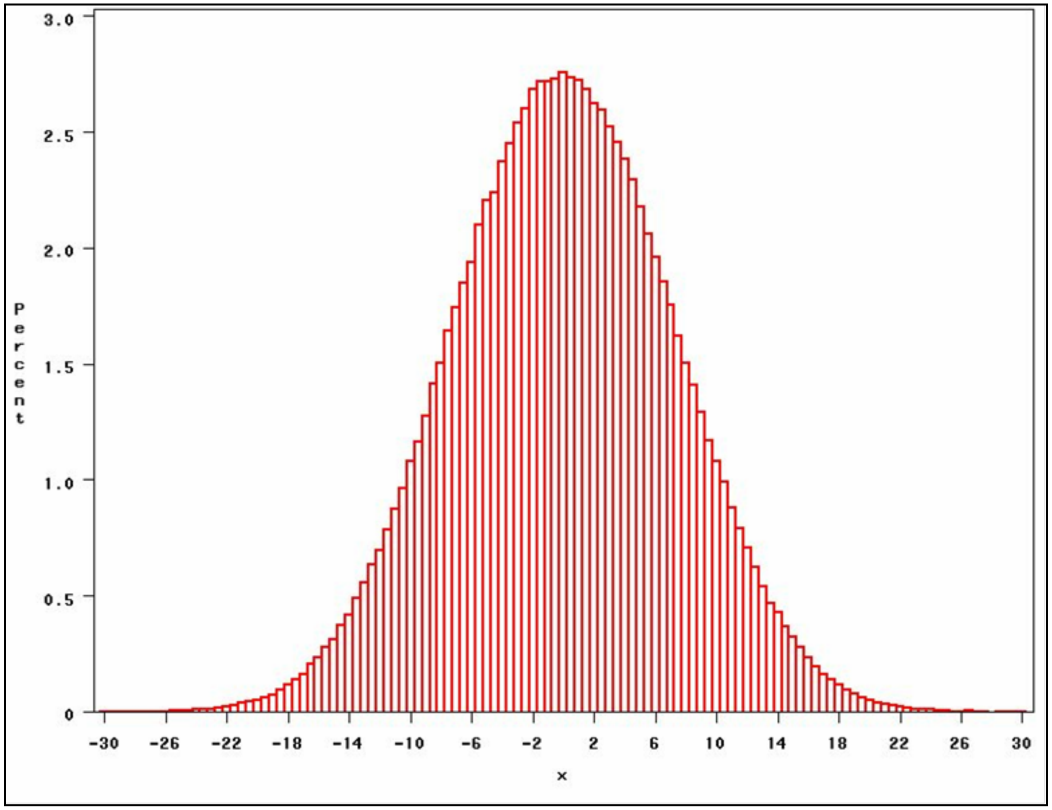
Consensus determines the scale of the search

(ES strategy parameters)



Trajectory of 1-D particle with fixed “bests”

Analyzing Particle Search



Previous bests constant

A million iterations, outliers trimmed

Histogram of points sampled

Q: What is the distribution of points that are tested by the particle?

Gaussian is a good guess. Mean of the distribution is halfway between the previous bests, standard deviation is relative to the difference between them.

“Bare Bones” particle swarm

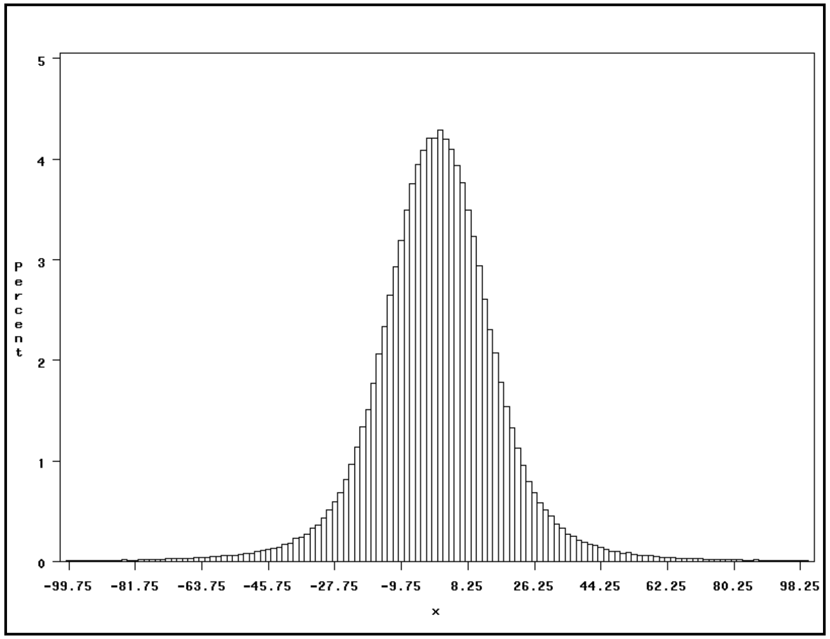
$$\boxed{x_i = G((p_i + p_g)/2, \mathrm{abs}(p_i - p_g))}$$

G(mean, s.d.) is Gaussian RNG

Simplified (!) – no system parameters to adjust

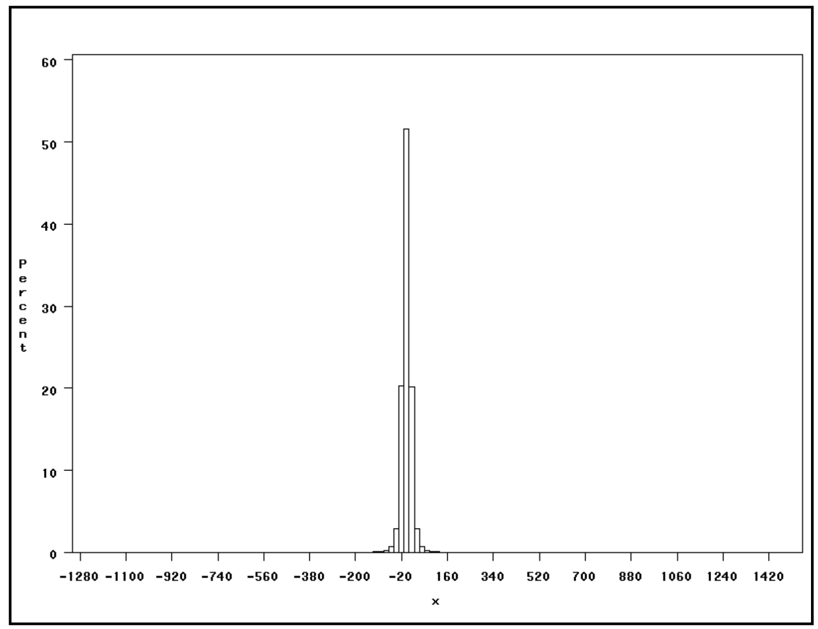
Works pretty well, but not as good as canonical.

Kurtosis in particle swarm



Tails trimmed

Peaked -- fat tails



Not trimmed

Empirical observations with p's held constant

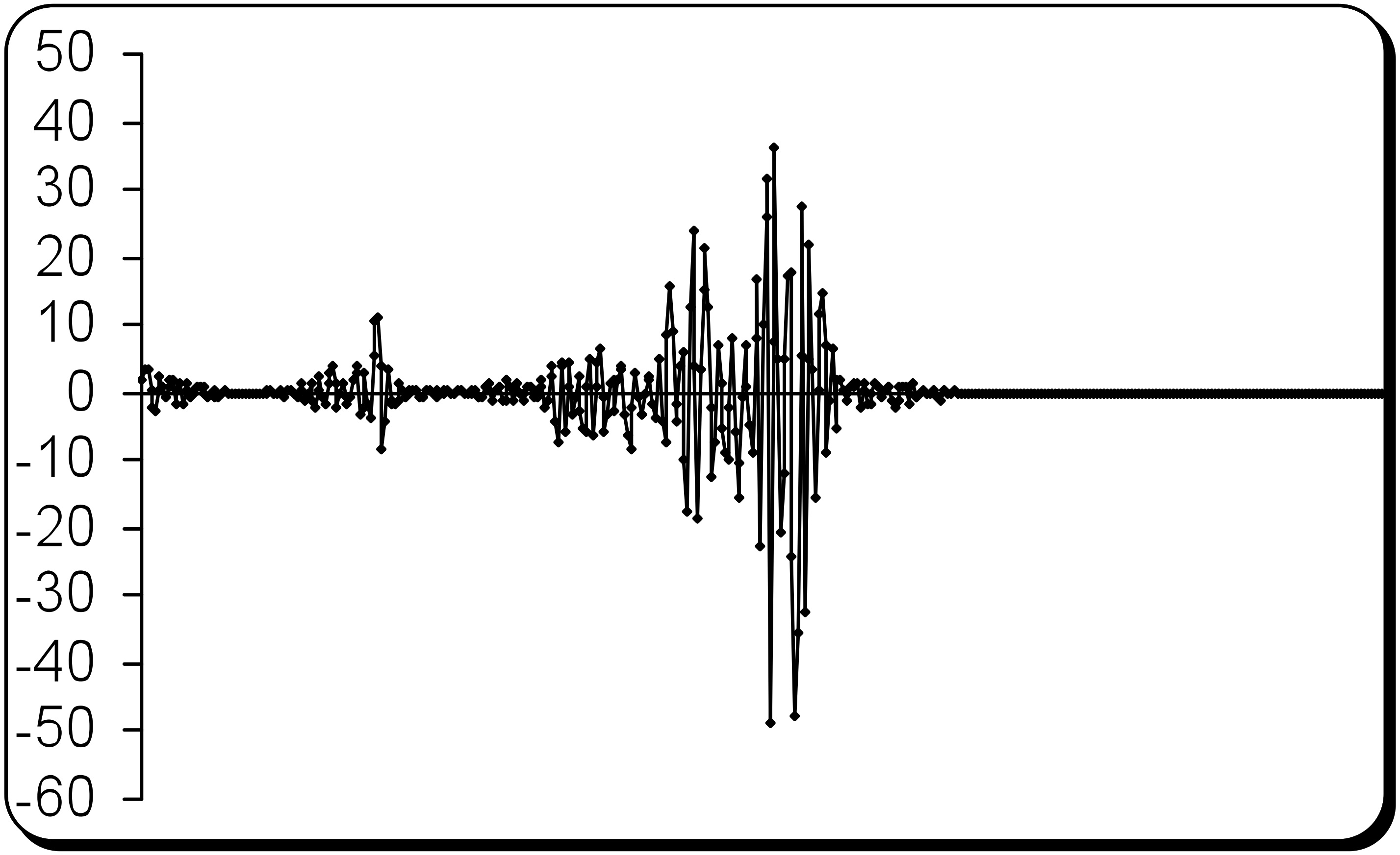
Kurtosis

High peak, fat tails

Mean moments of the canonical particle swarm algorithm with previous bests set at ± 20, varying the number of iterations.

Iterations	Mean	S.D.	Skew-ness	Kurtosis
1,000	0.0970	37.7303	-0.0617	8.008
3,000	0.0214	41.5281	0.0814	18.813
10,000	-0.0080	41.6614	-0.0679	40.494
100,000	0.0022	41.7229	0.2116	170.204
1,000,000	0.0080	41.3048	0.3808	342.986

Bursts of Outliers



"Volatility clustering" seems to typify the particle's trajectory

Adding Bursts of Outliers to Bare Bones PSO

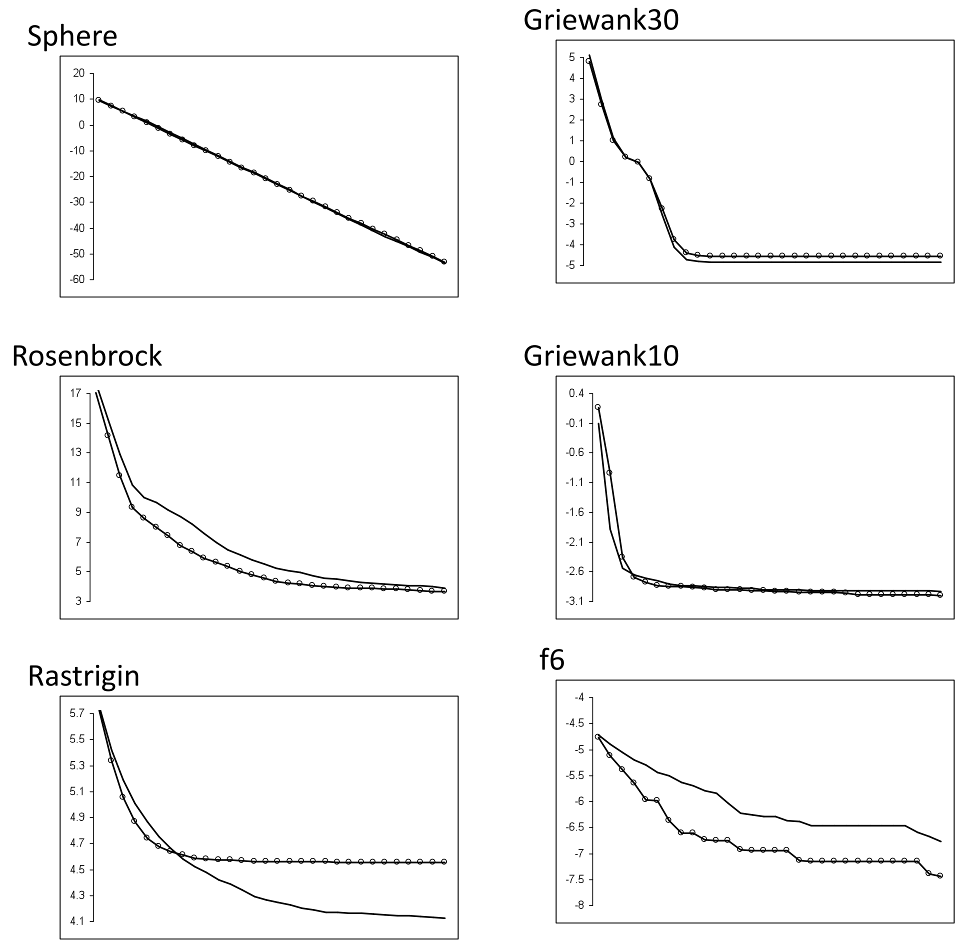
Center = $(p_{id} + p_{gd})/2$
SD = $|p_{id} - p_{gd}|$
x_{id} = G(0,1)

if Burst = 0 and U(0,1)< `PBurstStart` then
  Burst = U(0, `maxpower`)

Else If Burst > 0 and
  U(0,1)< `PBurstEnd` then
  Burst = 0
End If

If Burst > 0 then $x_{id} = x_{id}$ ^ Burst
x_{id} = Center + x_{id} * SD

(Bubbled line is canonical PS)



(Note that Bare Bones without bursts usually performs somewhat worse than canonical version)

Lévy Bare Bones – Richer and Blackwell

More points at very small distances and at very large distances from the mean

“The motivation is that the power law behaviour of the Lévy distribution at large step length ("fat tails") will induce exploration at any stage of the convergence, enabling escape from local minima. The Lévy PSO should reproduce the "‘Gaussian with Bursts"’ idea, but within a simpler and more intuitive scheme.”

Lévy performed better than Gaussian in this study.

Bursts support global search but are probably not the most effective approach

Richer., T., & Blackwell, T. M. (2006). The Lévy particle swarm. In *Proceedings of IEEE Congress on Evolutionary Computation, 3150–3157.*

Peña's Reduced Algorithm

Jorge Peña – discrete recombination for hardware implementation

for each dimension i **do**

$r = RAND()$

if $r = 0$ **then**

$k = left(j)$

else

$k = right(j)$

end if

$v_{id} = w \cdot v_{ji} + (p_{ki} - x_{ji}) + (p_{ji} - x_{ji})$

$v_{ji} \in (-V_{max}, V_{max})$

$x_{ji} = x_{ji} + v_{ji}$

end for

Coefficients approximated

One one-bit random number per dimension

(Peña, Upegui and Sanchez. Particle swarm optimization with discrete recombination: an online optimizer for evolvable hardware. Proceedings of the First NASA/ESA Conference on Adaptive Hardware and Systems (AHS) 2006, 163-170)

Bratton and Blackwell's Simplified Version

Bratton and Blackwell took the idea of bare-bones particles, analyzed the bursting behavior of a standard particle swarm, and developed a simplified algorithm. Uses neighbors to left and right.

$$\boxed{x_{id}^{t+1} = x_{id}^{t} + \phi(r_{id} - x_{id}^{t})}$$

where

$$\boxed{r_{id} = \eta_d p_{ld} + (1 - \eta_d) p_{rd}}$$

and Φ~1.2, η_d= U(0,1)

(Bratton, D. and Blackwell, T. (2008). A simplified recombinant PSO. *Journal of Artificial Evolution and Applications*, Article ID 654184.)

Poli et al's Simplification

$$v_{t+1} = \chi \left(v_t + \frac{\phi}{2}((g-p)u + p - x_t) \right)$$

where *u* is U(0,1)

"... SPSO is simpler to define, understand and control than the canonical PSO."

Performs "competitively" on test functions

(Poli, Bratton, Blackwell, & Kennedy. Theoretical derivation, analysis and empirical evaluation of a simpler particle swarm optimiser. Proceedings of 2007 IEEE Congress on Evolutionary Computation, 1955-1962.)

Peña: Particle Swarm With Additive Stochasticity

Following up on Poli et al.

General model:

$$x_{t+1} = x_t + w\left(x_t - x_{t-1}\right) + \alpha\left(q - x_t\right)$$

where *q* is a recombination operator

Uses various recombination operators, some borrowed from ES.

"From the point of view of the performance, and in the light of the presented empirical results no recombination operator seems to be superior to another. Recombination operators interact with parameters *w* and *Φ* for determining a PSO with specific sampling distribution characteristics that could be beneficial for some functions, but detrimental for others."

(Jorge Peña (2008). Theoretical and empirical study of particle swarms with additive stochasticity and different recombination operators. *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*, (GECCO), 2008, 95-102.)

Constriction and Inertia

Originally there was Vmax

Constriction coefficient

$$\begin{cases} v_{id}^{(t+1)} \leftarrow \chi\left(v_{id}^{t} + U(0,1)\left(\frac{\varphi}{2}\right)(p_{id} - x_{id}^{t}) + U(0,1)\left(\frac{\varphi}{2}\right)(p_{gd} - x_{id}^{t})\right), \\ x_{id}^{(t+1)} \leftarrow x_{id}^{t} + v_{id}^{(t+1)} \end{cases}$$

Inertia weight

$$\boldsymbol{v}_{k+1}^{i} = w_k \boldsymbol{v}_k^{i} + c_1 r_1 \left(\boldsymbol{p}_k^{i} - \boldsymbol{x}_k^{i}\right) + c_2 r_2 \left(\boldsymbol{p}_k^{g} - \boldsymbol{x}_k^{i}\right)$$

Inertia weight distributes χ across terms, removes parentheses

(M. Clerc and J. Kennedy (2002). The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation, 6(1):58–73.)*

Time-Decreasing Inertia Weight

$$\boldsymbol{v}_{k+1}^{i} = w_k \boldsymbol{v}_k^{i} + c_1 r_1 \left(\boldsymbol{p}_k^{i} - \boldsymbol{x}_k^{i}\right) + c_2 r_2 \left(\boldsymbol{p}_k^{g} - \boldsymbol{x}_k^{i}\right)$$

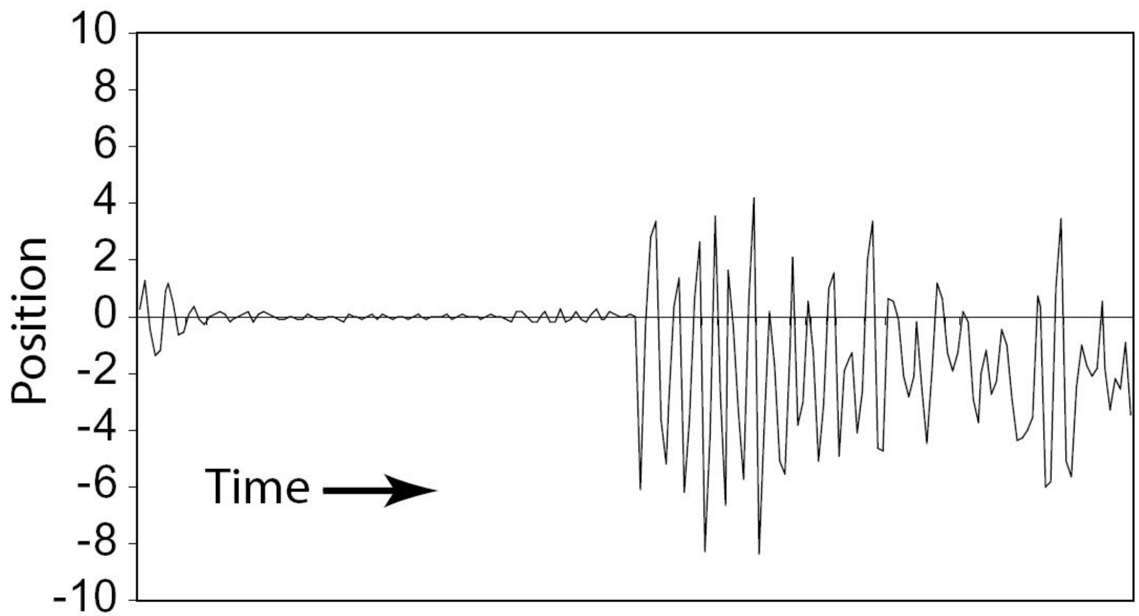
Reduces value of w_k over the iterations, typically from 0.9 to 0.4.

Supposedly induces “local” search

--You can’t go back

Recovery

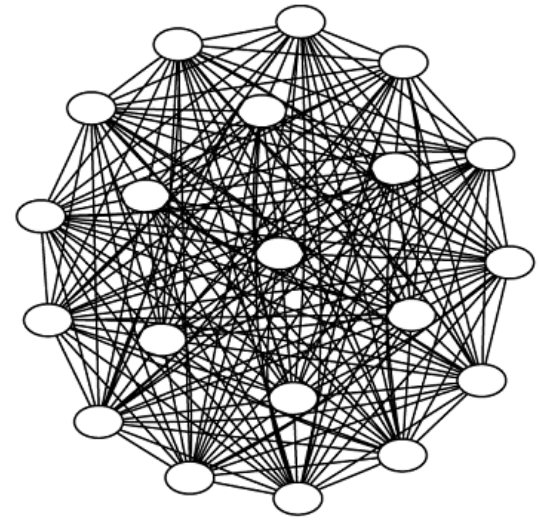
One-dimensional trajectory, new p_g found after 100 iterations



If, after some number of iterations, information comes to the particle about an optimum in a different region, the particle trajectory can expand and search between and beyond "here" and "there."

Time-decreasing inertia weight can't do this.

Gbest



"Fully connected" topology

Means the population's best solution informs all members of the population

All particles are influenced to search in the direction of the same point

Result may be premature convergence

Very many researchers complain about the "tendency of PSO to converge prematurely and get stuck in local optima" AND use Gbest topology

Plus it's uninteresting

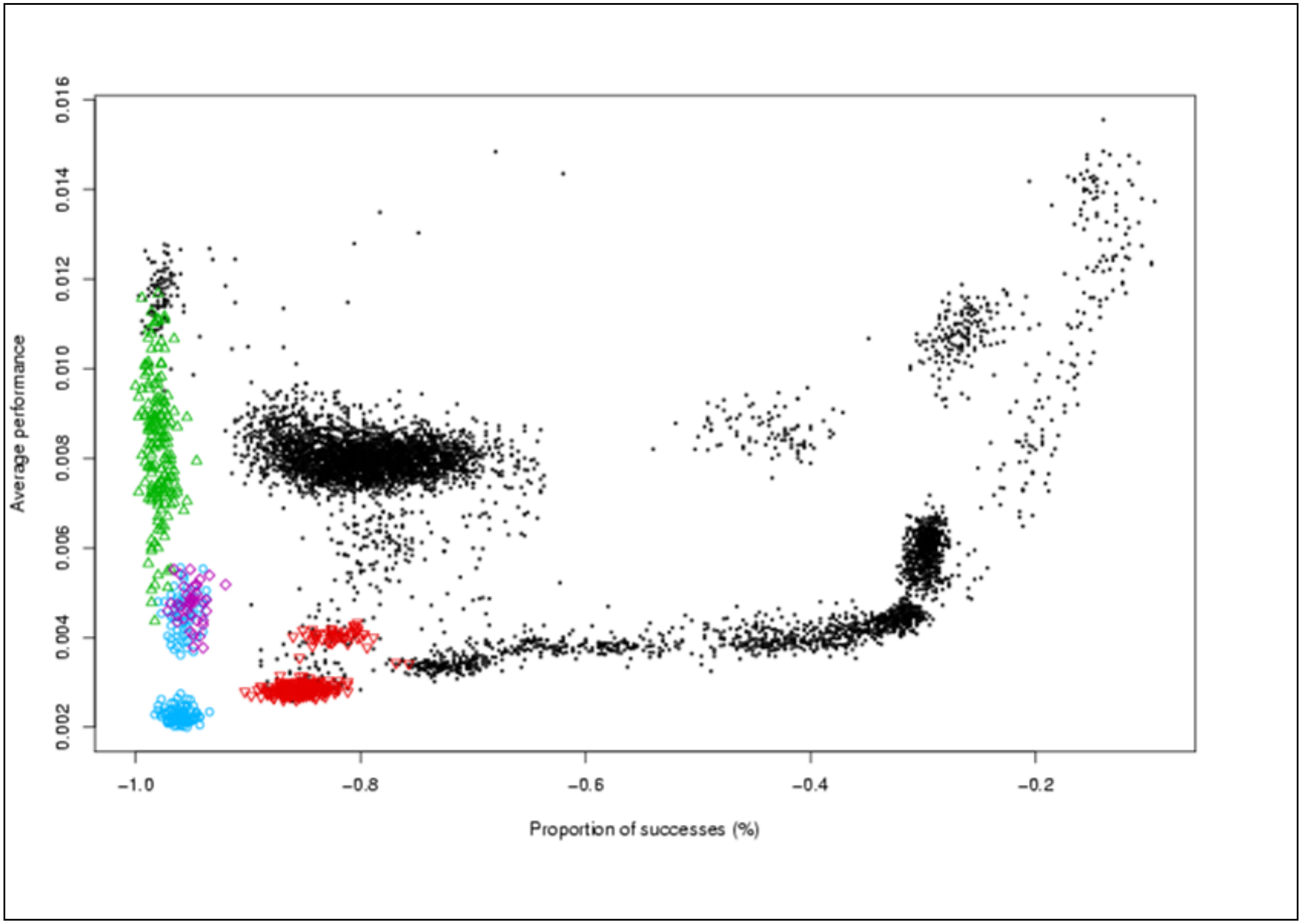
FIPS -- The "fully-informed" particle swarm (Rui Mendes)

$$
\begin{aligned}
v_{id}^{(t+1)} &\leftarrow \alpha v_{id}^{t} + \frac{\sum U(0,\beta)(p_{nbr_k d} - x_{id}^{t})}{K} \\
x_{id}^{(t+1)} &\leftarrow x_{id}^{t} + v_{id}^{(t+1)}
\end{aligned}
$$

Note that p_i is not a source of influence in FIPS.
Doesn't select best neighbor but uses them all.
Orbits around the mean of neighborhood bests.
This version is more dependent on topology.

(Gbest for example is *very* slow and tends to converge on initial average)

Mendes: Two Measures of Performance



Color and shape indicate parameters of the social network – degree, clustering, etc.

Key to Graph

- **cyan circles** Topologies with average degree in the interval (4, 4.25). The elements of this class display the best average performances while maintaining a high proportion of successes.

- **magenta diamonds** Topologies whose average degree belongs to the interval (3, 3.25) and clustering coefficient in the interval (0.7, 0.9). This class is not as good as the previous one.

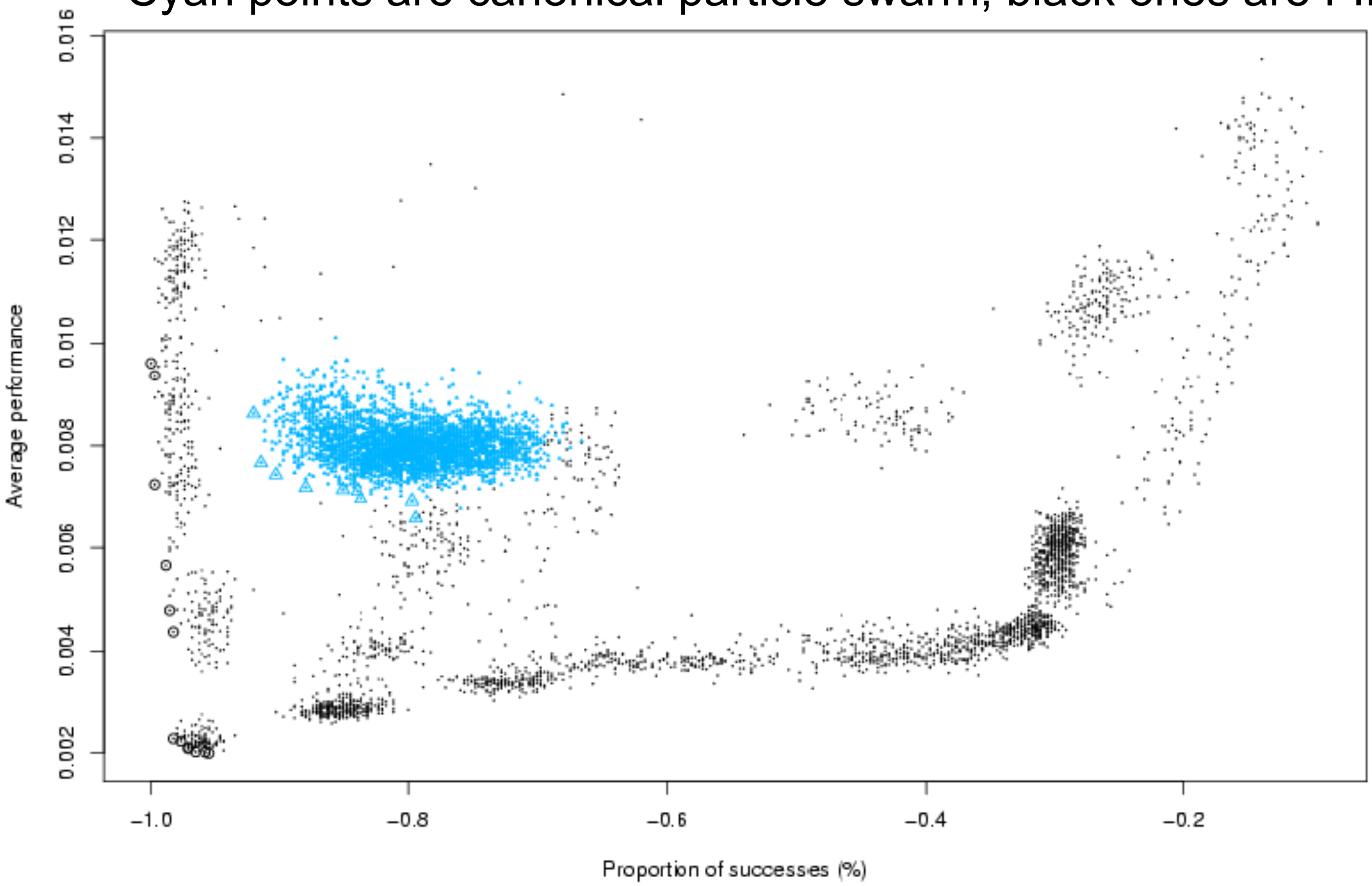
- **green triangles** Topologies with average degree in the interval (3, 3.25) and clustering coefficient in the interval (0.1, 0.6). This class has a worse average performance but contains some individuals with the highest proportion of successes.

- **red inverted triangles** Topologies with average degree in the interval (5, 6) and clustering coefficient in the interval (0.025, 0.4). This class has other high quality average of successes while maintaining proportions of successes of at least 80%.

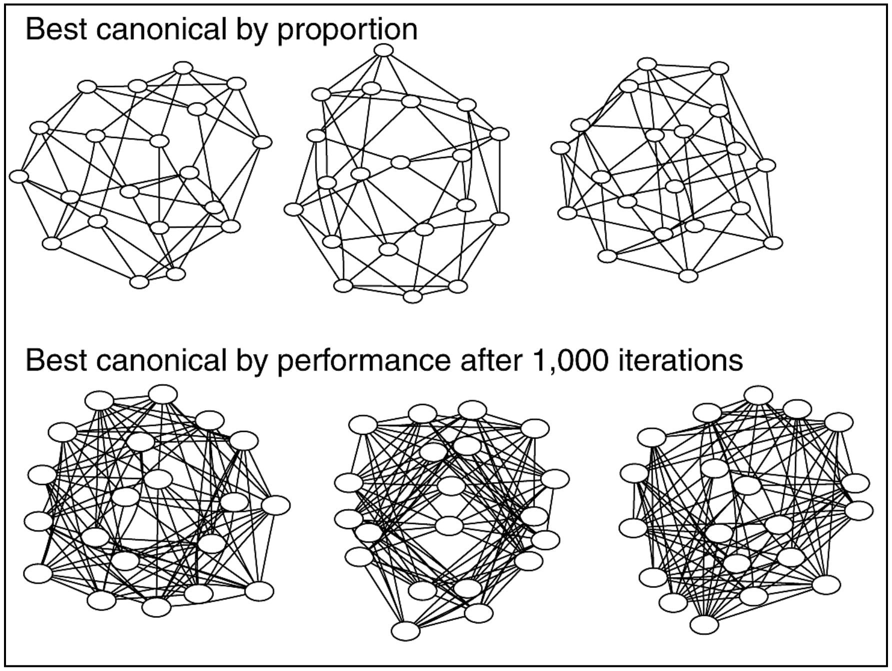
- **black dots** All the topologies that do not belong to any of the above classes are displayed as dots to have a measure of comparison.

Mode of Interaction and Network Topology

Cyan points are canonical particle swarm, black ones are FIPS

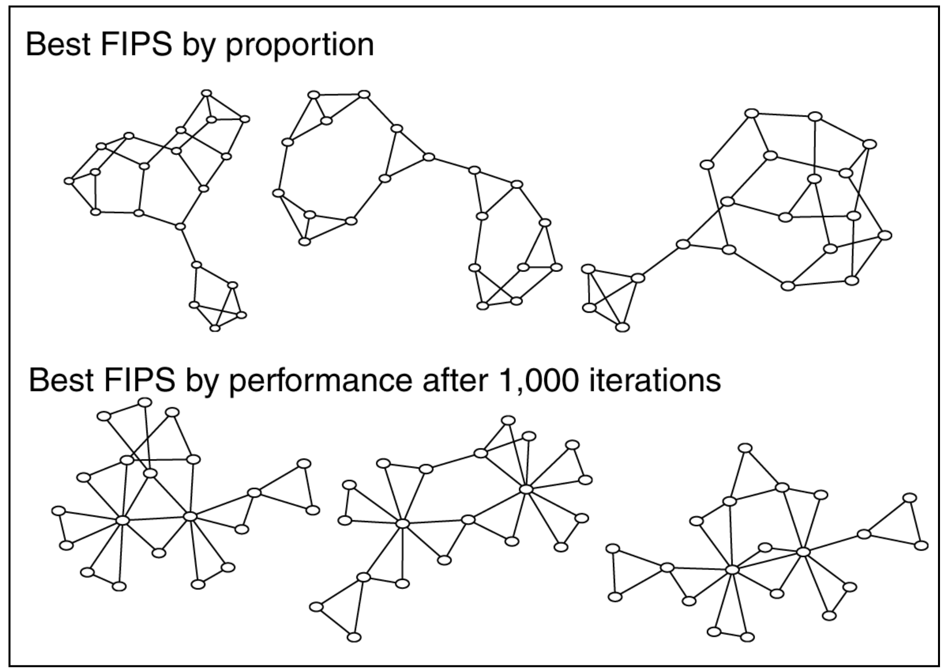


Mode of Interaction and Network Topology



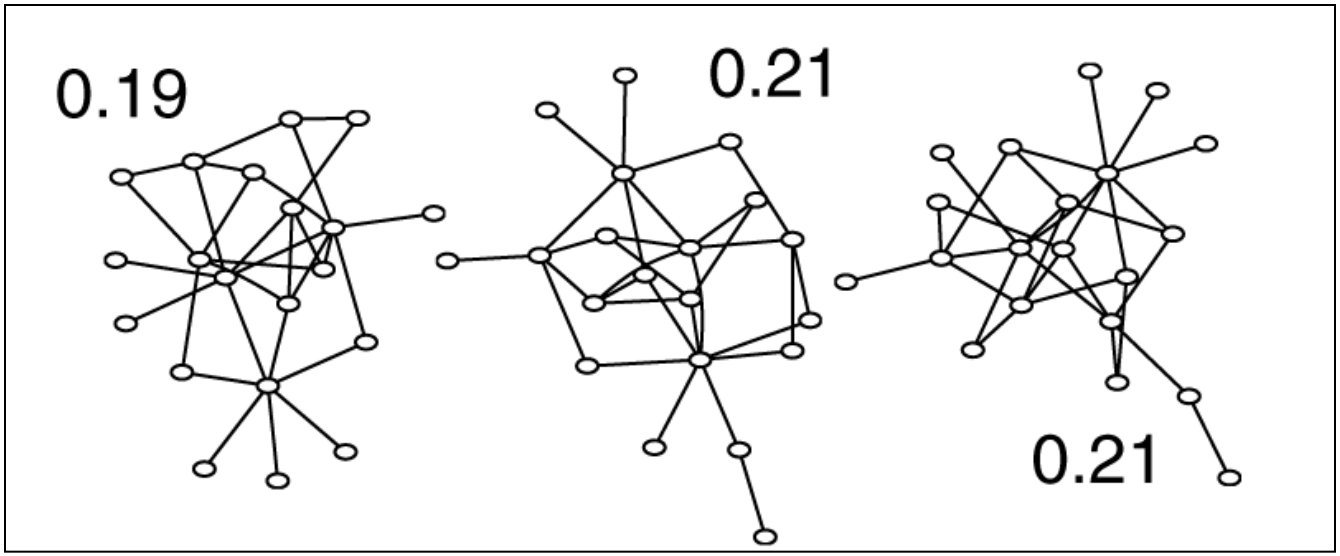
Best-neighbor versions

FIPS versions



Worst FIPS Sociometries

and Proportions Successful



Interaction Mode

Traditional: best-neighbor interaction

- Compare the best performances of all topological neighbors
- Receive influence from the best one and from the self

Fully Informed Particle Swarm (FIPS)

- Stochastic average of all neighbors' previous bests
- Not the self

Other possibilities

- Pick one or *n* at random
- Probabilistic interaction
- Fuzzy neighbors
- Weighted links

The possibilities are endless and have not been well explored

ES recombination (e.g., Peña's study)

Look at human society for examples – pedagogy vs. norms, for instance

Deconstructing Velocity

$$
\begin{aligned}
v_{id}^{(t+1)} &\leftarrow \alpha v_{id}^{t} + \\
& U(0,1)(\frac{\beta}{2})(p_{id} - x_{id}^{t}) + \\
& U(0,1)(\frac{\beta}{2})(p_{gd} - x_{id}^{t}) \\
x_{id}^{(t+1)} &\leftarrow x_{id}^{t} + v_{id}^{(t+1)}
\end{aligned}
$$

$$
\begin{aligned}
x_{id}^{(t+1)} \leftarrow x_{id}^{t} & \\
& + \alpha\left(x_{id}^{t} - x_{id}^{(t-1)}\right) \\
& + U(0,1)\left(\frac{\beta}{2}\right)\left(p_{id} - x_{id}^{t}\right) \\
& + U(0,1)\left(\frac{\beta}{2}\right)\left(p_{gd} - x_{id}^{t}\right)
\end{aligned}
$$

Canonical particle swarm can be written in one formula

Generalization

We can generalize the canonical and FIPS versions:

$$
\begin{aligned}
x_{id}^{(t+1)} \quad \leftarrow \quad & x_{id}^{t} + \\
& \alpha(x_{id}^{t} - x_{id}^{(t-1)}) + \\
& \sum \left(U(0,1)\frac{\beta}{K}(p_{nbr_k d} - x_{id}^{t}) \right)
\end{aligned}
$$

The only difference is how you choose the sources of influence.

Verbal Representation

NEW POSITION =
CURRENT POSITION +
PERSISTENCE +
SOCIAL INFLUENCE

$$
\begin{aligned}
x_{id}^{(t+1)} \quad \leftarrow \quad & x_{id}^{t} + \\
& \alpha(x_{id}^{t} - x_{id}^{(t-1)}) + \\
& \sum \left(U(0,1)\frac{\beta}{K}(p_{nbr_k d} - x_{id}^{t}) \right)
\end{aligned}
$$

Huge Particle Swarms

Google MapReduce: "a programming model and computation platform for parallel computing. It allows simple programs to benefit from advanced mechanisms for communication, load balancing, and fault tolerance."

Scales to 256 processors on moderately difficult problems and tolerates node failures.

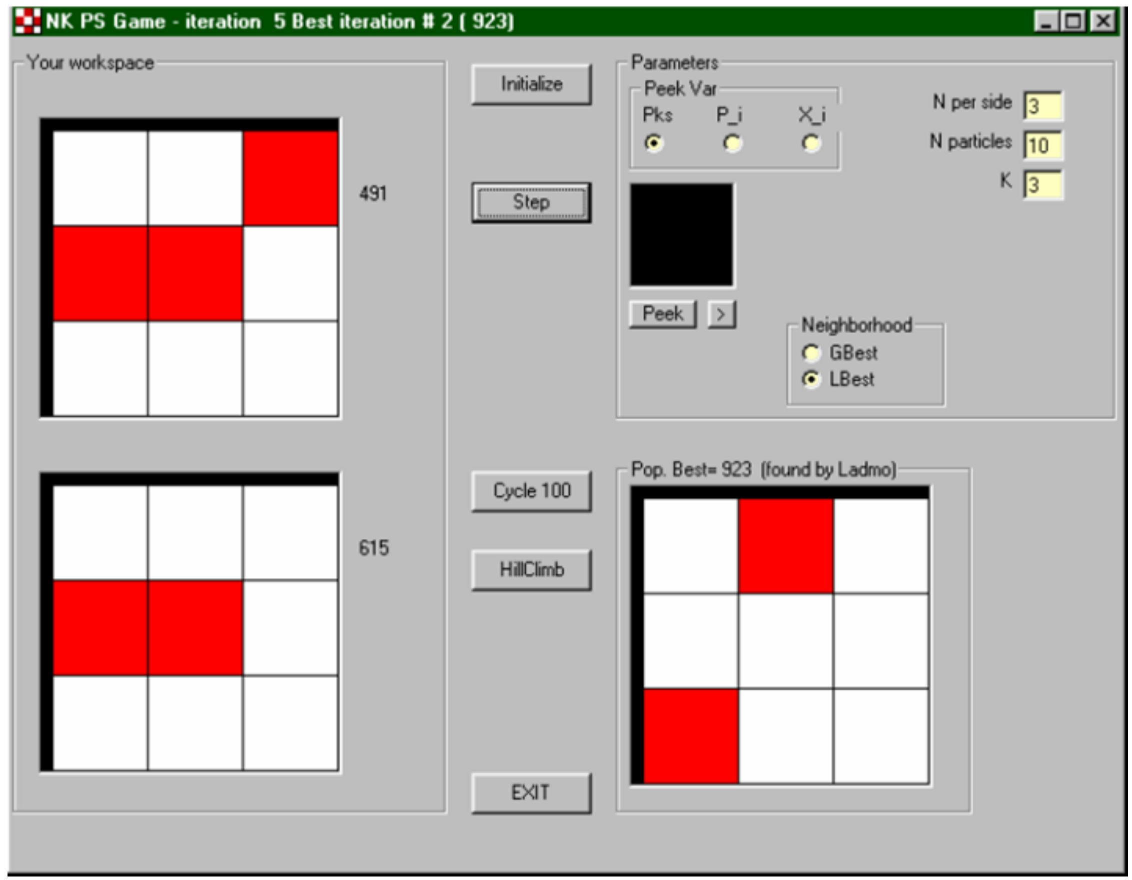
Swarm = 1,000 particles

"Makes no assumption about whether the sociometry is static or dynamic"

McNabb, Monson, Seppi (2007). Parallel PSO using MapReduce. Proceedings of IEEE Congress on Evolutionary Computation.

"Exteriorized" Particle Swarm

User is "a particle"
Could be used for training, education, group problem solving
Mix humans and digital entities in various proportions



Also note musical improvisations, etc

Growth medium

Future Focus of the Evolving Paradigm

The paradigm has been stuck in some local optima, seems to be escaping them

Language and sophisticated particle processing (esp. in huge swarms)

Simplification and finding essentials

What is important about the population topology?

Opening the algorithm to integrate digital and human participants

Taking ideas from social psychology, human interaction

Applying ideas to organizations, paradigms

Evolving a holistic focus, looking at the swarm as an entity

Send me a note:

kennedy.jim@gmail.com